

The Genetic Activity Profile Database

by Michael D. Waters,* H. Frank Stack,† Neil E. Garrett,† and Marcus A. Jackson†

A graphic approach termed a Genetic Activity Profile (GAP) has been developed to display a matrix of data on the genetic and related effects of selected chemical agents. The profiles provide a visual overview of the quantitative (doses) and qualitative (test results) data for each chemical. Either the lowest effective dose (LED) or highest ineffective dose (HID) is recorded for each agent and bioassay. Up to 200 different test systems are represented across the GAP. Bioassay systems are organized according to the phylogeny of the test organisms and the end points of genetic activity. The methodology for the production and evaluation of GAPs has been developed in collaboration with the International Agency for Research on Cancer. Data on individual chemicals have been compiled by IARC and by the U.S. Environmental Protection Agency. Data are available on 299 compounds selected from volumes 1-50 of the *IARC Monographs* and on 115 compounds identified as Superfund Priority Substances. Software to display the GAPs on an IBM-compatible personal computer is available from the authors.

Structurally similar compounds frequently display qualitatively and quantitatively similar GAPs. By examining the patterns of GAPs of pairs and groups of chemicals, it is possible to make more informed decisions regarding the selection of test batteries to be used in evaluating chemical analogs. GAPs have provided useful data for the development of weight-of-evidence hazard ranking schemes. Also, some knowledge of the potential genetic activity of complex environmental mixtures may be gained from assessing the GAPs of component chemicals. The fundamental techniques and computer programs devised for the GAP database may be used to develop similar databases in other disciplines.

Introduction

Data derived from short-term tests are usually interpreted according to the phylogenetic category of the test and the end point detected. Commonly studied end points include DNA damage, gene mutation, sister chromatid exchange, micronuclei, chromosomal aberrations, aneuploidy, and cell transformation. Few short-term bioassays monitor more than one or two of these end points. Therefore, data from a variety of short-term tests are required to properly define the response profile of a given chemical agent.

Garrett et al. (1) developed a technique for presenting the quantitative genetic toxicology data for a chemical compound as a bar graph (genetic activity profile) in which test systems (identified by three-letter code words) are displayed along the X-axis, and values corresponding to the doses employed in the tests are shown on the Y-axis. The total data available from up to 200 different short-term bioassays for a compound are thus presented in a standardized format that allows rapid visualization of the genetic (or related) effects induced. The technique facilitates qualitative as well as quantitative assessments of genetic toxicity. Current procedures for preparing and evaluating Genetic

Activity Profiles (GAPs) are described by Waters et al. (2) in the context of their use by the International Agency for Research on Cancer.

Methodology

The data set for a given chemical, consisting of a discrete set of tests and the doses required to induce responses in those tests, are presented in a bar graph illustrated in Figure 1. The bars (profile lines) originating on the X-axis represent the tests plotted in either a phylogenetic or end point sequence. A three-letter code is used to identify the test system represented by each bar. Values on the Y-axis are the logarithmically transformed lowest effective doses (LED) and highest ineffective doses (HID) tested. The term "dose," as used in this report, does not take into consideration length of treatment or exposure and may therefore be considered synonymous with concentration. The doses or concentrations used for all *in vitro* tests were converted to micrograms per milliliter and those for *in vivo* tests to milligrams per kilogram body weight per day. Because dose units are plotted on a log scale, differences in molecular weights of compounds do not greatly influence comparisons of their GAPs.

Profile-line height (the magnitude of each bar) is a function of the LED or HID, which is associated with the characteristics of each individual test system, such as population size, cell-cycle kinetics, and metabolic competence. Thus, the detection limit of each test system is different, and across a given GAP, responses will vary substantially. No attempt is made to adjust or relate responses in one test system to those of another.

*Genetic Toxicology Division, Health Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711.

†Environmental Health Research and Testing, Inc., Research Triangle Park, NC 27709.

Address reprint requests to M. D. Waters, Genetic Toxicology Division, Health Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711.

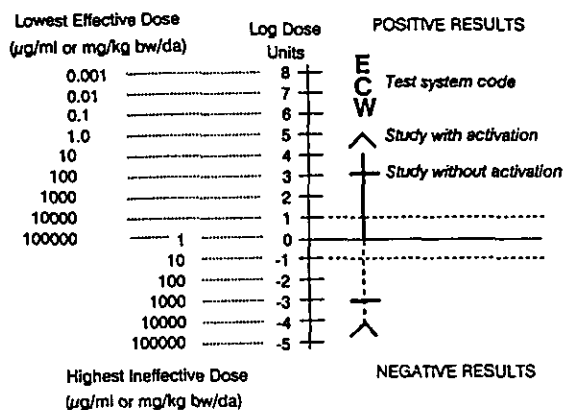


FIGURE 1. A schematic representation of a genetic activity profile showing four studies for the test ECW (two positive and two negative). The average logarithmic dose unit of the majority call is indicated by a solid vertical bar. A dashed vertical bar indicates conflicting test results for the study. Note in cases where there are an equal number of positive and negative studies as here, the majority call is positive.

Line heights are derived as follows: For negative test results, the highest dose tested without excessive toxicity is defined as the HID. If there is evidence of extreme toxicity, the next lower dose is used. A single dose tested yielding a negative result is considered equivalent to the HID. For positive results, the LED is recorded. If the original data have been analyzed statistically by the author, the dose recorded is that at which the response was significant ($p < 0.05$). If the data were not analyzed statistically, the dose required to produce an effect is estimated as follows: When a dose-related positive response is observed with two or more doses, the lower of the doses is taken as the LED; a single dose resulting in a positive response is considered equivalent to the LED.

To accommodate both positive and negative responses on a continuous scale, doses are transformed logarithmically so that effective (LED) and ineffective (HID) doses are represented by positive and negative numbers, respectively. The logarithmic dose unit (LDU_{ij}) for a given test system i and chemical j is represented by the expressions:

$$LDU_{ij} = -\log_{10}(\text{dose}), \text{ for HID values; } LDU < 0$$

$$LDU_{ij} = 5 - \log_{10}(\text{dose}), \text{ for LED values; } LDU > 0$$

These simple relationships define a dose range of 0 to -5 logarithmic units for ineffective doses (1–100,000 $\mu\text{g/mL}$ or mg/kg body weight) and 0 to +9 logarithmic units for effective doses (100,000–0.0001 $\mu\text{g/mL}$ or mg/kg body weight). A scale illustrating the LDU values is shown in Figure 1. Negative responses at doses less than 1 $\mu\text{g/mL}$ (mg/kg body weight) are set equal to 1. Effectively, an LED value $> 100,000$ or an HID value < 1 produces an $LDU = 0$; no quantitative information is gained from such extreme values. Levels of log dose units between 1 and -1 define a “zone of uncertainty” in which positive results are reported at very high doses (10,000–100,000 $\mu\text{g/mL}$ or mg/kg body weight), and negative results are reported at relatively low dose levels (1–10 $\mu\text{g/mL}$ or mg/kg body weight).

All dose values are plotted for each assay using either a bar (–) for results obtained in the absence of an exogenous metabolic system or a caret (∧) for those obtained in the presence of an ex-

ogenous metabolic system. When all results for a given assay are either positive or negative, the geometric mean of the responses is plotted as a solid line; when conflicting data are reported for the same assay (i.e., both positive and negative results), the majority data are shown with a solid line and the minority data with a dashed line (drawn to the extreme response). In the few cases where the numbers of positive and negative results are equal, the solid line is drawn in the positive direction, and the negative response is indicated with a dashed line, drawn from the origin to the extreme negative LDU.

The three-letter code words representing the commonly used tests were originally defined by the Gene-Tox Program of the U.S. Environmental Protection Agency (EPA) (3,4). These codes have been systematically redefined and expanded in a manner that should facilitate inclusion of additional tests in the future (2).

Evaluation of Genetic and Related Effects

The International Agency for Research on Cancer (IARC) has employed the GAP methodology (2) in evaluating genetic and related effects of suspected human carcinogens in *IARC Monographs*, supplement 6 (5) and in volumes 36, 39, 41, 44 (13–16) and 46–50 (6–10). Table 1 illustrates the procedure currently employed by the IARC as it relates to the GAP database.

Nesnow (personal communication) has recently completed a PC D-Base version of the data on carcinogenicity contained in *IARC Monographs*, supplement 7 (17) and *IARC Monographs*, volumes 43 (18), 44 (16), 45 (19), and 46–49 (6–9). Most of these agents are included in the GAP database [derived from *IARC Monographs*, supplement 6 (5) and volumes 46–50 (6–10)]. Therefore, these two databases can be used to examine retrospectively the usefulness of short-term tests for the prediction of carcinogenicity and the relationship between specific genetic end points or assays and carcinogenicity.

Personal Computer Version of the GAP Database, Version 3.0

Copies of software for IBM-compatible personal computers to display and search GAPs are available from the authors. Computer programs require the following minimum configuration: PC using Intel 8086 chip (PC XT) with 640 kb memory, a hard disk drive, an enhanced graphics card (EGA or VGA), a high-resolution color monitor, and DOS version 3.2 or higher.

Table 1. Current process of chemical selection and review of data on genetic and related effects used to prepare genetic activity profiles for *IARC Monographs*.

Before the <i>IARC Monograph</i> working group meeting
1. IARC selects compounds and completes literature search
2. Working subgroup members review literature for individual compound(s)
3. Subgroup member prepares written summary, summary table, and data listing, drafts are sent to IARC
During the <i>IARC Monograph</i> group meeting
4. Genetic and related effects subgroup convenes
5. Data summaries and tests are verified for each compound
6. Final drafts are submitted to working group for review
7. Data tables and summaries are sent to EPA/Genetic Toxicology Division for dose verification
8. Genetic activity profiles and data listings are prepared and sent to IARC

Generally, an Intel 80286 (PC AT) computer is preferred because data processing and graphics display are faster than with the 8086 computer. Optional devices used for data and graphic output include a line printer and plotter. Alternatively, a laser printer or equivalent can be used to print the HPGL plotter files using additional software.

The GAP software is distributed on three double-sided, double-density, 5.25-inch floppy disks; the program disk, the data disk, and the GAP bibliography. Executable programs are archived on the program disk and are compiled from programs written in Turbo Pascal. During installation of the programs, the necessary directory and subdirectories are created. The installed programs and data use approximately 1.2 Mb of disk space.

The bibliography of the GAP data is also in an archived file. The file requires 0.7 Mb of disk space; however, the bibliography is not necessary to operate the GAP programs, and it may be deleted and reinstalled as needed.

The data disk consists of two data sets, IARC and EPA. The IARC data set contains data on 299 agents published in supplement 6 (5) and in volumes 46–50 of the *IARC Monographs* (6–10). The EPA data set contains data on 115 agents assembled for the Genetic Toxicology Division of the U.S. Environmental Protection Agency (11). A list of the individual projects included in each data set may be viewed using the GAP computer program. A data subdirectory is provided for users to enter their own data.

Main Program Menu

The main program menu of GAP version 3.0 offers the following selections: agents, profiles, data listings, modify data, short citations, and additional information. "Agents" provides options to list the available projects, CAS numbers, and agent names. Another menu allows ordering of the list by any of the three options. "Profiles" provides graphic display of the short-term test data on selected agents. A menu is used to select the sequence of test codes, either in phylogenetic order of organisms (i.e., prokaryotes, lower eukaryotes, etc.) or in test end point order (i.e., DNA damage, gene mutation, etc.). Individual test codes may be examined to determine the source citations by using the GAP program zoom-in features.

"Listings" produces a listing of the data in either phylogenetic or end point order and may be directed to the PC screen, to a printer, or to a data file. "Modify data" is used to add, change, or delete agents or test results (test codes, results, doses, and reference numbers).

"Short Citations" permits searching the literature citation information for approximately 6000 short citations contained in the GAP database. The citation information includes the citation number (LITNR), the Environmental Mutagen Information Center (EMIC) accession number, and a short citation (consisting of the last names of up to three authors, the first page number and the year of the publication). The citation information may be searched by author or by EMIC number to determine if a citation is present. Short citations also may be added to the file and are automatically assigned citation numbers.

"Additional information" includes three-letter test code definitions, the scale of log-dose units used in the profiles, information on the dose conversions, and tables listing projects for both the EPA and IARC data sets.

Some Applications of the GAP Database

Comparative Evaluation of Genetic Activity Profiles Using Computer-Based Profile-Matching Techniques

Where an adequate number of the same tests have been used to evaluate two or more chemicals, it is possible to use the main-frame computer to select matching pairs of GAPs. This computer-based pairwise matching process may be extended to all chemicals in the database. The pilot applications of this procedure to EPA databases on known or suspected human carcinogens (1) and on pesticide chemicals (12) have demonstrated that structurally similar compounds frequently display qualitatively and quantitatively similar profiles of genetic activity. This implies that the GAP database should be of considerable utility in structure-activity relationship investigations and in test battery selection (20).

By examining the patterns of GAPs of pairs and groups of chemicals, it is possible to make more informed decisions regarding the selection of test batteries to be used in the subsequent evaluation of structurally similar chemicals. The approach draws on all information within the database and may be linked to computer systems that model the molecular properties of the chemicals under evaluation (21). This comparative information can enhance our understanding of the relationships between genetic and related activity in short-term tests and molecular properties of structurally related chemicals and thus contribute to our knowledge of the mechanisms of complex processes such as carcinogenesis.

Testing and Evaluating Complex Mixtures

A recent application of GAPs is in testing and evaluating complex mixtures (22). Some knowledge of the potential genetic activity of a complex environmental mixture may be gained from assessing the genetic activity of its component chemicals. This requires information on the chemical components and composition of the mixture. For example, the Atmospheric Chemical Compound database developed by Graedel et al. (23) contains information on chemical structures, properties, detection methods, and sources of chemicals found in ambient air. The GAP database provides a computer-generated graphic representation of genetic bioassay data as a function of dose. Using the two databases, information on the quantity of an individual chemical present within a mixture may be related to the quantity (LED) of the chemical required to demonstrate a positive response in one or more genetic bioassays. Quantitative information on the carcinogenic potency of each individual compound (TD_{50} value) may also be related to the quantity present in the mixture or mixture fraction. In turn, the quantity of the chemical in the complex mixture to which humans are exposed may be estimated and used to calculate the percent human exposure dose/rodent potency dose (HERP) for the chemical (24–26). Using an additivity assumption, for example, an estimate of potential carcinogenic hazard for the mixture may be calculated based on the HERP indices for the known chemical components. This conceptual approach is limited by the relatively small number of chemicals identified in complex mixtures for which genetic toxicology and animal cancer data exist.

Weight-of-Evidence Ranking Schemes

Committee 1 of the International Commission for Protection Against Environmental Mutagens and Carcinogens (ICPEMC) has for several years been involved in the development of a computer-based methodology to assess the evidence from short-term genetic tests that a chemical is a mutagen (27). The evaluative approach selected by ICPEMC Committee 1 is based on a "weighted test" scoring system that provides a relative ranking of genotoxic potential. Input data for this ranking methodology have been obtained from the GAP database described above (28). The results of the application of the Committee 1 ranking scheme are to be compared by ICPEMC to results obtained by applying the carcinogenicity ranking scheme of Nesnow (29,30).

Development of Other Databases

The fundamental techniques and computer programs devised for the GAP database may be used to develop similar databases in genetic toxicology and in other disciplines. Dearfield et al. (31) have described the application of the GAP methodology to the database being constructed by the EPA Office of Pesticide Programs. Kavlock et al. (32) have successfully used the approach and modified computer programs to assemble graphic activity profiles and corresponding data listings for several developmental toxicants.

Future Directions

A useful application of the GAP database in the future will involve computer-based profile-matching techniques with weight-of-evidence ranking schemes to create a subset of chemicals that act similarly, i.e., have similar GAPs. Correlative structure-activity approaches can then be used more effectively to identify the substructural elements of chemicals that are responsible for particular biological responses so as to suggest biologically plausible mechanisms of action (33).

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. The authors would like to acknowledge the following contributions to the GAP database: computer programming, Paul H.M. Lohman and Walter J.A. Lohman, University of Leiden, Leiden, The Netherlands; data input, Ann Brady, Environmental Health Research and Testing, Inc.; Beth Owens, John Wassom, and Elizabeth Von Halle, Oak Ridge National Laboratories; and data evaluation, Douglas McGregor, Harri Vainio, and Linda Shuker, International Agency for Research on Cancer, Lyon, France.

REFERENCES

- Garrett, N. E., Stack, H. F., Gross, M. R., and Waters, M. D. An analysis of the spectra of genetic activity produced by known or suspected human carcinogens. *Mutat. Res.* 134: 89-111 (1984).
- Waters, M. D., Stack, H. F., Brady, A. L., Lohman, P. H. M., Haroun, L., and Vainio, H. Use of computerized data listings and activity profiles of genetic and related effects in the review of 195 compounds. *Mutat. Res.* 205: 295-312 (1988).
- Waters, M. D. The GENE-TOX program. In: *Banbury Report 2* (A. W. Hsie, J. P. O'Neill, and V. K. McElheny, Eds.), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1979, pp. 449-467.
- Waters, M. D., and Auletta, A. The GENE-TOX program: genetic activity evaluation. *J. Chem. Inf. Comput. Sci.* 21: 35-38 (1981).
- IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans—Genetic and Related Effects: An Updating of Selected IARC Monographs from Volumes 1 to 42, Supplement 6. International Agency for Research on Cancer, Lyon, France, 1987.
- IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 46. Diesel and Gasoline Engine Exhausts and Some Nitroarenes. International Agency for Research on Cancer, Lyon, France, 1989.
- IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 47. Some Organic Solvents, Resin Monomers and Related Compounds, Pigments and Occupational Exposures in Paint Manufacture and Painting. International Agency for Research on Cancer, Lyon, France, 1989.
- IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 48. Some Flame Retardants and Textile Chemicals, and Exposures in the Textile Manufacturing Industry. International Agency for Research on Cancer, Lyon, France, 1990.
- IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 49. Chromium, Nickel, and Welding. International Agency for Research on Cancer, Lyon, France, 1990.
- IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 50. Pharmaceutical Drugs. International Agency for Research on Cancer, Lyon, France, 1990.
- Waters, M. D., Stack, H. F., and Jackson, M. A. Genetic activity profiles on hazardous substances listed under SARA, Sect. 110—The First 100 Superfund Priority Substances, Internal Report. U.S. Environmental Protection Agency, Washington, DC, 1990.
- Garrett, N. E., Stack, H. F., and Waters, M. D. Evaluation of the genetic activity profiles of 65 pesticides. *Mutat. Res.* 168: 301-325 (1986).
- IARC. IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans, Vol. 36. Allyl Compounds, Aldehydes, Epoxides and Peroxides. International Agency for Research on Cancer, Lyon, France, 1985.
- IARC. IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans, Vol. 39. Some Chemicals Used in Plastics and Elastomers. International Agency for Research on Cancer, Lyon, France, 1985.
- IARC. IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans, Vol. 41. Some Halogenated Hydrocarbons and Pesticide Exposures. International Agency for Research on Cancer, Lyon, France, 1986.
- IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 44. Alcohol Drinking. International Agency for Research on Cancer, Lyon, France, 1988.
- IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Overall Evaluation of Carcinogenicity: An Updating of Selected IARC Monographs from Volumes 1 to 42, Supplement 7. International Agency for Research on Cancer, Lyon, France, 1987.
- IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 43. Man-Made Mineral Fibres and Radon. International Agency for Research on Cancer, Lyon, France, 1988.
- IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 45. Occupational Exposures in Petroleum Refining; Crude Oil and Major Petroleum Fuels. International Agency for Research on Cancer, Lyon, France, 1989.
- Waters, M. D., Stack, H. F., Rabinowitz, J. R., and Garrett, N. E. Genetic activity profiles and pattern recognition in test battery selection. *Mutat. Res.* 205: 119-138 (1988).
- Richard, A. M., Rabinowitz, J. R., and Waters, M. D. Strategies for the use of computational SAR methods in assessing genotoxicity. *Mutat. Res.* 221: 181-196 (1989).
- Waters, M. D., Claxton, L. D., Stack, H. F., Brady, A. L., and Graedel, T. E. Genetic activity profiles in the testing and evaluation of chemical mixtures. *Teratog. Carcinog. Mutagen.* 10: 147-164 (1990).
- Graedel, T. E., Hawkins, D. T., and Claxton, L. D. *Atmospheric Chemical Compounds: Sources, Occurrence, and Bioassay*. Academic Press, Orlando, FL, 1986.
- Gold, L. S., Backman, G., Hooper, K., and Peto, R. Ranking the potential carcinogenic hazards to workers from exposures to chemicals that are tumorigenic in rodents. *Environ. Health Perspect.* 76: 211-220 (1987).
- Gold, L. S., Sawyer, C. B., Magaw, R., Backman, G. M., de Veciana, M., Levinson, R., Hooper, N. K., Havender, W. R., Bernstein, L., Peto, R., Pike, M. C., and Ames, B. N. A carcinogenic potency database of the standardized results of animal bioassays. *Environ. Health Perspect.* 58: 9-319 (1984).

26. Gold, L. S., Slone, T. H., Manley, N. B., Backman, G. M., Garfinkel, G., Hudes, E. S., Rohrbach, L., and Ames, B. N. The carcinogenic potency database: analyses of 4000 chronic animal cancer experiments published in the general literature and by the U.S. National Cancer Institute/National Toxicology Program. *Environ. Health Perspect.* 96: 11-15 (1991).
27. Lohman, P. H. M., Mendelsohn, M. L., Moore, D. H., II, Waters, M. D., and Brusick, D. J. The assembly and analysis of short-term genotoxicity test data—An ICPEMC Committee 1 working paper. In: *Mutation and the Environment, Part D* (M. L. Mendelsohn and R. J. Albertini, Eds.), Wiley-Liss, 1990, pp. 283-294.
28. Brusick, D. A proposed method for assembly and interpretation of short-term test data. *Environ. Health Perspect.* 96: 101-111 (1991).
29. Nesnow, S. A multi-factor ranking scheme for comparing the carcinogenic activity of chemicals. *Mutat. Res.* 134: 89-111 (1984).
30. Nesnow, S. A multifactor ranking scheme for comparing the carcinogenic activity of chemicals. *Environ. Health Perspect.* 96: 17-21 (1991).
31. Dearfield, K. L., Quest, J. A., Whiting, R. J., Stack, H. F., and Waters, M. D. Characteristics of the U.S. EPA's Office of Pesticide Programs' toxicity information databases. *Environ. Health Perspect.* 96: 53-56 (1991).
32. Kavlock, R. J., Greene, J. A., Kimmel, G. L., Morrissey, R. E., Owens, E., Rogers, J. M., Sadler, T. W., Stack, H. F., Waters, M. D., and Welsch, F. Activity profiles of developmental toxicity: design considerations and pilot implementation. *Teratology* 43:159-185 (1991).
33. Waters, M. D., Richard, A. M., Rabinowitz, J. R., Stack, H. F., Garrett, N. E., Lohman, P. H. M., and Rosenkranz, H. S. Structure-activity relationships—computerized systems. In: *Proceedings of the Scientific Group on Methodologies for the Safety Evaluation of Chemicals (SGOMSEC)-8 Workshop on Cross-Species Differences in DNA Damage and its Consequences*. Research Triangle Park, NC, 1990, in press.